Data Management 101 for Earth Scientists

Creating Documentation and Metadata

Nancy J. Hoebelheinrich

Knowledge Motifs LLC

Mapping sensible data relationships

Roadmap

- What is Metadata?
- How is it relevant to Data Management?
- Distinction b/w Metadata and Documentation
- Metadata Types & Functions
- Best Practices for Metadata Creation

Definition of Metadata

Information to let you & others find, understand, and use your data both now and in the future

descriptions

descriptors aka keywords

documentation

Relevance to Data Management

- **Documentation** for My Latest Research Project

- **To Data Manager:**

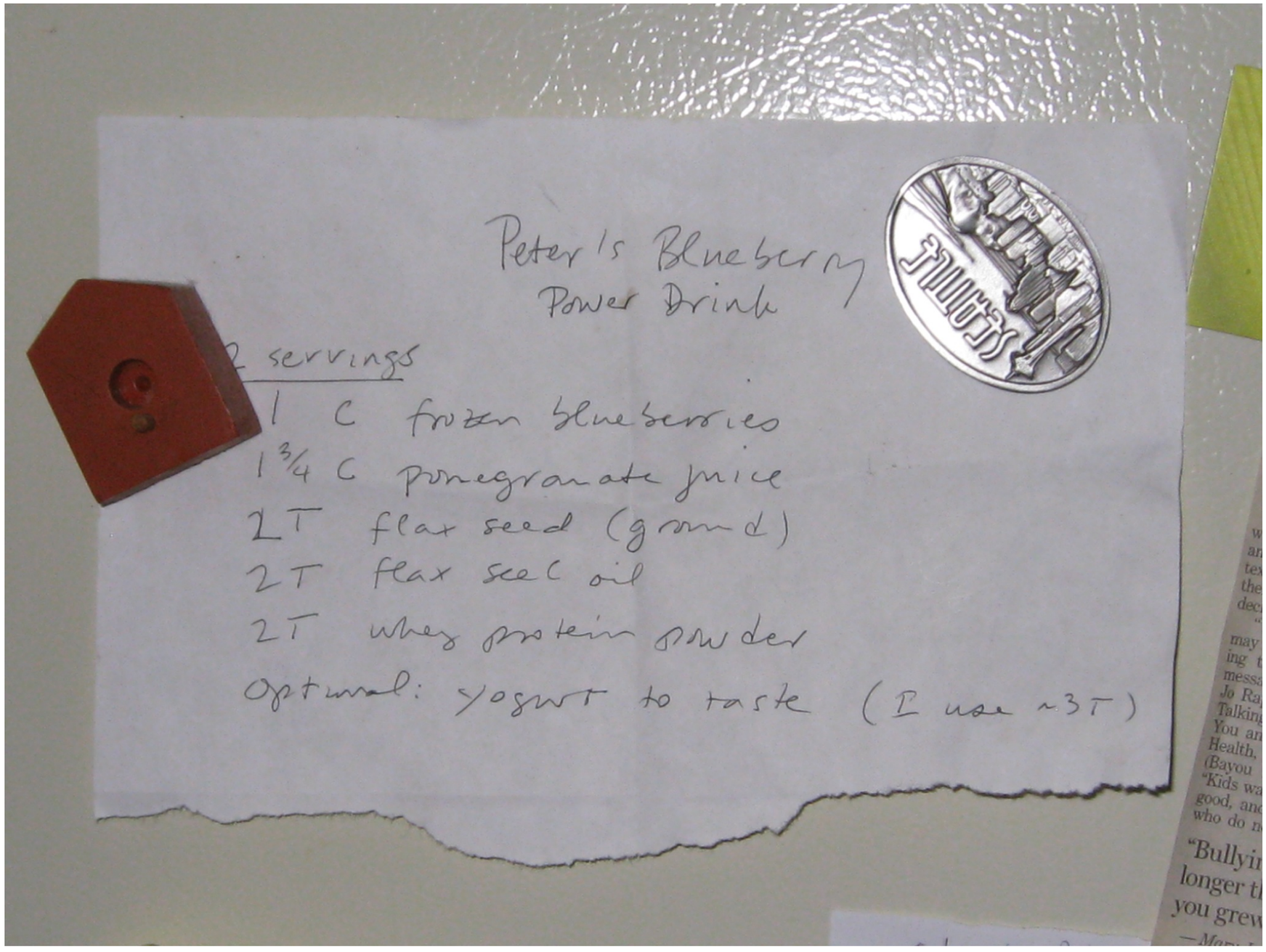
- ***Don't worry, the connections are all there [– in my head!]***

Relevance to Data Management

- **Documentation** for My Latest Research Project

- **To Data Manager:**

- ***See, here's the primary algorithm I used…oops, where's the process?***

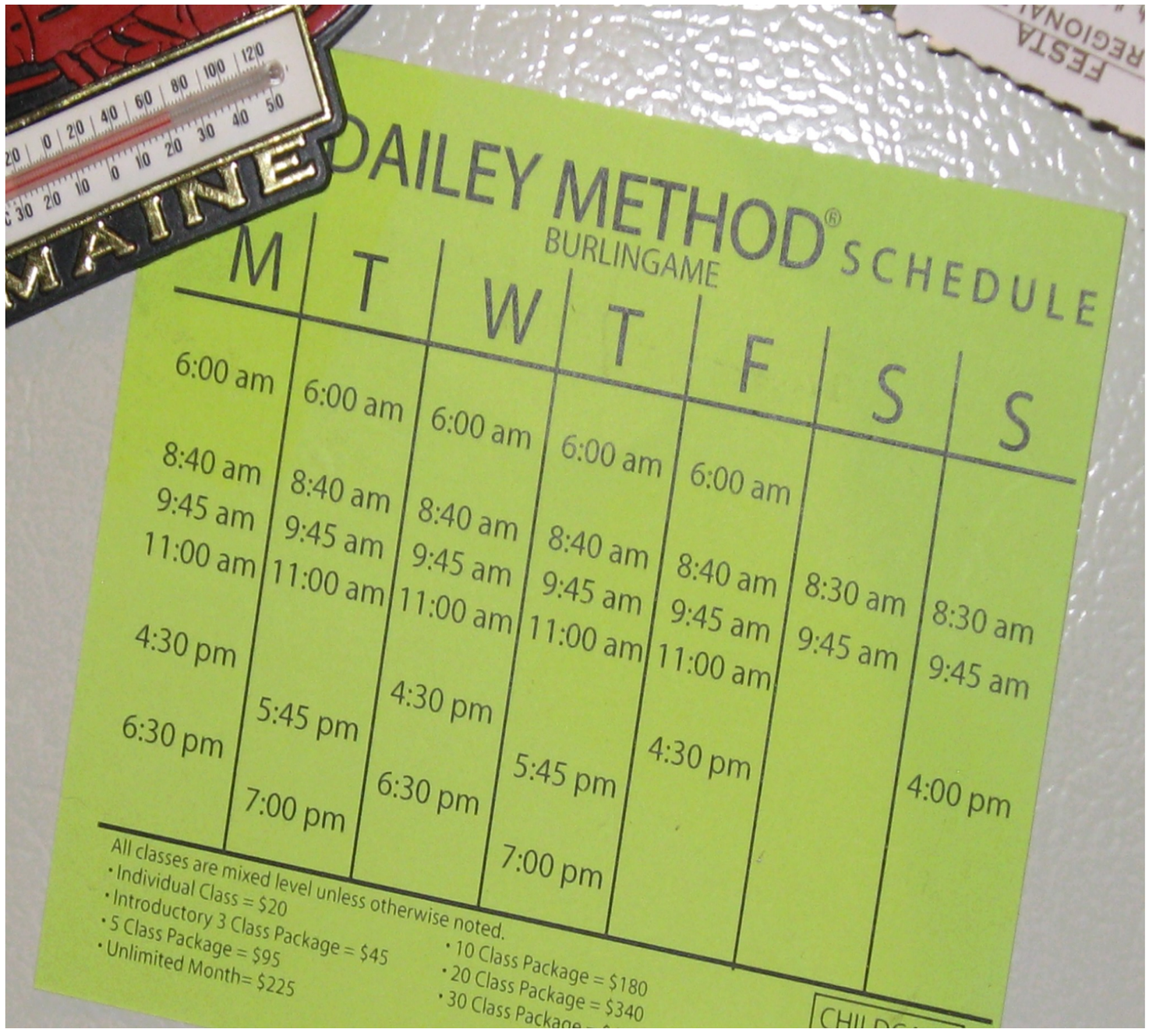


Relevance to Data Management

- **Documentation** for My Latest Research Project

- **To Data Manager:**

- ***Here's the schedule we used to gather the data – although some months it was a little different…***



Relevance to Data Management

Documentation
for My Latest
Research Project

To Data Manager:

Here's the research methodology we used – more or less…

Relevance to Data Management

- **Documentation** for My Latest Research Project

- **To Data Manager:**

- ***Oh, and here's the team – our PI wasn't there for the photo, so his placeholder is the guy with the mustache on the stick… And the project manager has the long ears… what was her name???***

You get the idea … so, what's needed?

- **Negotiated** b/w you *& your team (data producers), data developers* and the data archive
- **Created / collected** as part of workflows associated with **lifecycle** of the data
- Based on **key questions, e.g.,** :
  - How will I and others find my data?
  - What will future scientists need to know to understand and re-use the data?
  - What will data (product) developers need to create data services and products that use my data?
  - What does the data archive team need from the data producer team (as the ones who know the data best) in order to preserve it for the long term?

The 20-Year Rule (NRC 1991)

- The metadata accompanying a data set should be written for a user 20 years into the future--*what does that investigator need to know to use the data?*

- Prepare the data and metadata / documentation for a user who is unfamiliar with the details of your project, methods, and observations

- In short: the WHO, WHAT, WHEN, WHERE, HOW, WHY about your data

Distinction b/w “metadata” & “documentation”

- Metadata:
  - **Question & Answer Game**
  - Using **standards** (other people’s questions based on community agreement),
    - the data **archive asks** you & your project team questions
    - **you** & your project team (try to) **answer** them **as best you can**

- Documentation:
  - Sometimes called “**content**” specifications
  - Addresses key question of **<u>what</u>**
  - **Representation** will change from discipline to discipline and project to project (i.e., what form the documentation will take)

There may be *overlap* b/w these two depending upon who is speaking

How can you help?

Framework Questions:

- Besides me, who's going to care?
  - Sponsor mandates to archive
  - Specific requirements from sponsor
    - e.g., NASA, NOAA, USGS
  - Data archive requirements & desirements
    - Negotiated & documented in Submission Information Package (OAIS SIP)
  - Future scientists who want to use/re-use your data!!

- What kind of metadata & documentation will they need?
  - Look for existing formulae for decision making, e.g.,
    - NOAA National Climatic Data Center's Climate Data Record Maturity Matrix; factors include *software readiness, existence / state of metadata & (other) documentation, utility of data, validity of product (based on certainty estimates), desire for / restrictions upon public access*

Specific ?'s from different perspectives:

When submitting data:

- Why was the data was created?
- What limitations, if any, do the data have?
- What does the data mean?
- Who should be cited if someone publishes something that utilized your data?

When receiving data:

- What are the data gaps?
- What processes were used for creating the current data?
- Are there any fees associated with the data?
- In what scale were the data created?
- What do the values in the tables mean?
- What software do I need in order to read the data?
- What projection is the data in?
- Can I give this data to someone else?

Metadata Types and Functions

- **Provenance & Context**
  - Physical & conceptual environment needed to understand data
  - Important for use & re-use

- **Discovery** including:
  - Citation to enable proper credit, authority & identification
  - Access & use restrictions
- **Preservation / Archiving**
  - Facilitating utility of data over time
- **Project Documentation**
  - Accumulation of important facts, guidelines, explanations about the project

Provenance & Context

- "Don't touch that! You don't know where it's been."
- In law, the chain of custody follows a piece of evidence from collection through presentation and helps document its **authenticity** and **reliability**
- For the user community to confidently determine the **suitability** of your data for specific applications, you must adequately describe "where" it has been.

More about Provenance & Context

- Part of the **WHO**, **WHAT**, **WHEN**, **WHERE**, **WHY** and **HOW** details that apply to your data during its lifecycle –
  - from *collection*
  - through *processing*
  - to *final analysis*

- Comparing & combining data
- Citing sources
- Identifying what's needed to understand your data
- Process steps
- When & how to document
- Discipline specific example

Comparing and combining data

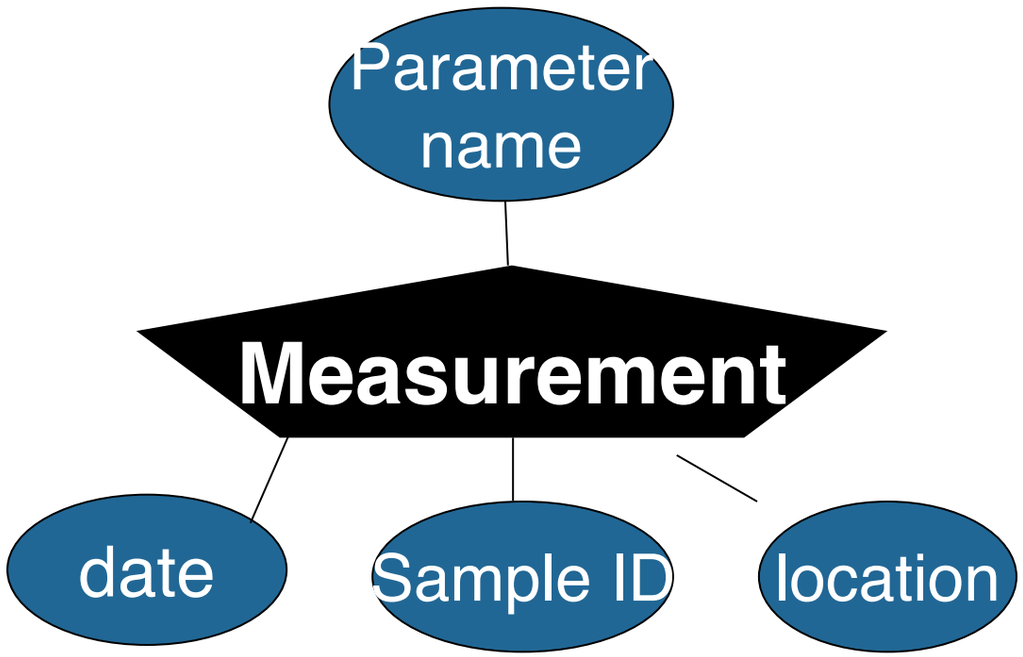
- Purposes:
  - When selecting amongst available data, the user must determine which data sets most closely **align** with the intended purpose
  - Understanding the reason data were collected in the first place lends insight into possible **biases**
- Advantages to combining data sets and tools from heterogeneous sources:
  - *expands* coverage
  - *extends* content
  - *provides* validity checks

Why cite data sources?

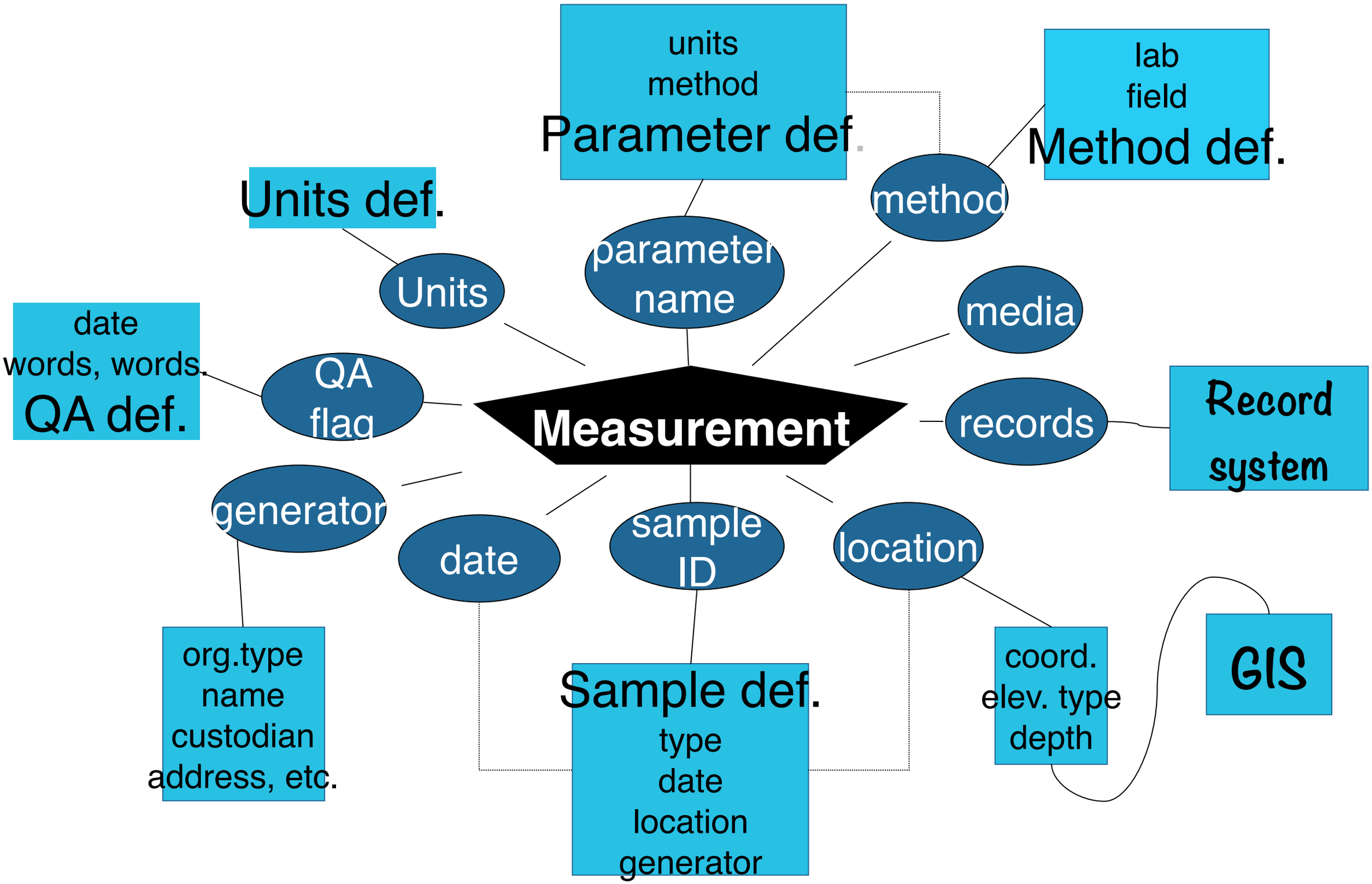
- To give credit where credit is due
- Simply using existing data from an archive or portal does not guarantee repeatable results.
  - Data may be updated or corrected
  - Issues with data delivery systems could affect the data received
- To ensure your results are repeatable, or that potential differences are accounted for, it's best to cite not only the source and its accompanying metadata record, but also
  - The dataset version or similar designator
  - The specific date(s) on which you acquired the data
  - The specific URL, request process or other method(s) through which you acquired the data
  - Any special subsetting logic or cross-reference information used to restrict the data received

What's needed to *Understand* my data?

The details of the data ….



Down to the nitty-gritty?



Why describe process steps?

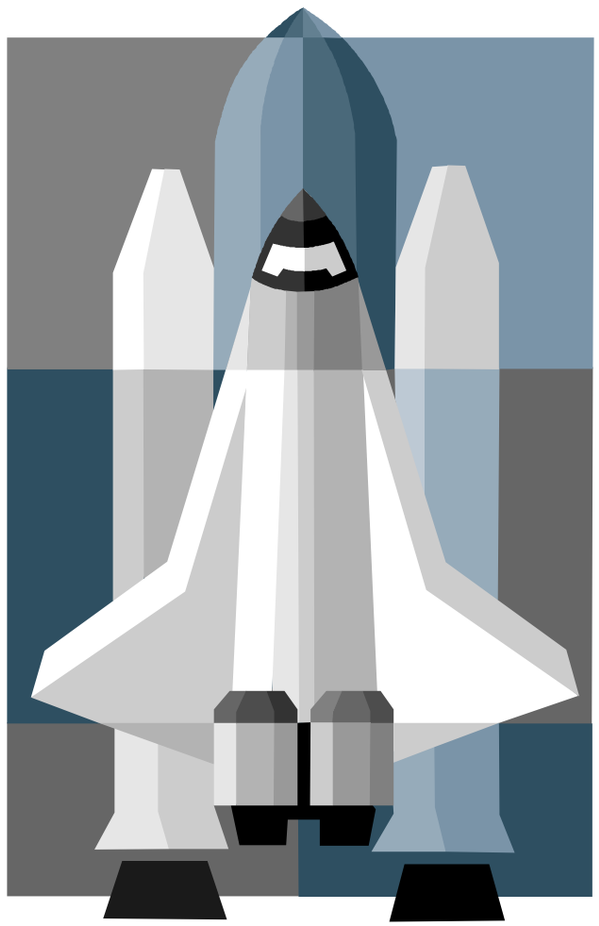
- Each time data are transformed, unintentional changes may also be introduced
- Documenting each processing step helps isolate changes and the potential for errors
- Such documentation may be general, or quite specific. More detail is usually better than less, so when in doubt, spell it out.
- If standard algorithms, libraries or other tools are used, they should be clearly referenced. In the case of libraries, GIS shape files and the like, be sure to document the source, version and date

Where and how to document

- Standards-based metadata records (ISO 19115-2, FGDC, and others) include areas in which various aspects of provenance and lineage may be documented, including
  - Data Quality
  - Source
  - Process steps
  - External citations and references
- Do not let a lack of familiarity with standards put you off!
  - It is relatively easy to get assistance in creating standards-based metadata records from a comprehensive set of details captured during your research
  - It is very difficult to recreate details that were not recorded fully

Provenance & Context

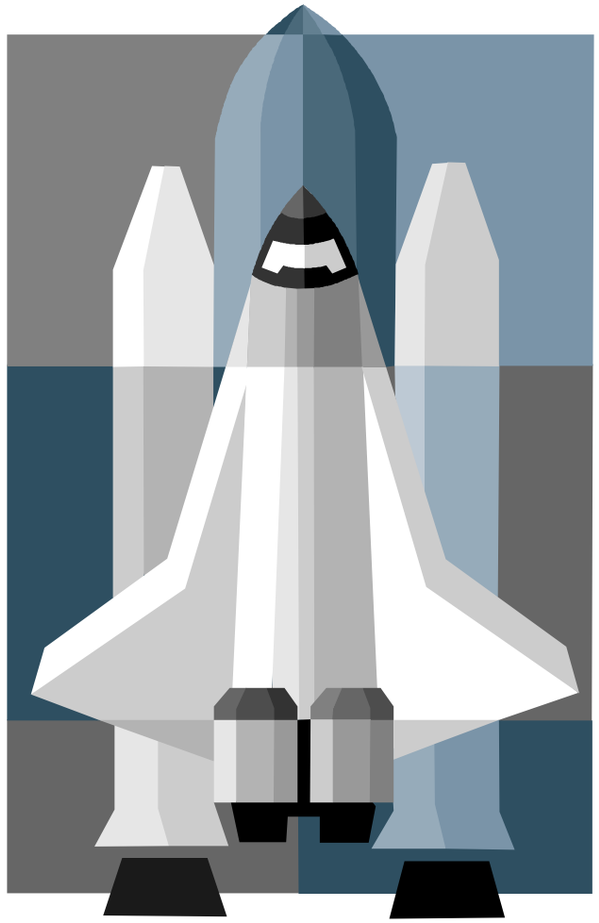
A Discipline Specific Example



- NASA Earth Science Data Preservation Content Specification (published Nov 2011)
  - Focused upon earth science data resulting from NASA's missions
  - Includes reference to a separate Metadata Requirements document for NASA Earth Science data products
  - Categorizes content into 8 areas:
    - Preflight/Pre-Operations Calibration, Science Data Products, Science Data Product Documentation, Mission Data Calibration, Science Data Product Software, Science Data Product Algorithm Input, Science Data Product Validation & Science Data Software Tools
    - Describes **rationale** for each, see example

Provenance & Context

A Discipline Specific Example



- **Preflight/Pre-Operations Calibration**
- **- 3.1.2 Preflight/Pre-operational Calibration Data**
- **Item Description**: **Numeric (digital data) files of Instrument/sensor characteristics**
  - including pre-flight or pre-operational performance measurements (e.g., spectral response, instrument geometric calibration (geo-location offsets), noise characteristics, etc.).
- **Rationale**: Measurements made before deploying instruments in space (or *in situ*) will help establish a baseline

Discovery & Citation Metadata

- Why is it needed?
  - To convey to the user the nature and content of the data
  - To help the user determine if a particular data set meets the users' needs
  - To provide the information necessary to create a citation for the data set
  - Used to "feed" data catalogs / libraries, information portals, citation systems
  - Can be used to describe or trigger restriction on access to data, and on use of the data
- What is it?
  - Descriptive information that allows users to find and assess the utility of data
    - Can be applied at / describe different levels, i.e., "collection", data set, and item level
    - Will usually contain less information than other types of metadata (i.e. use metadata, granule metadata)
    - Contains a set of quantitative and/or qualitative measurements and distinguishes that set from other similar measurement sets
    - Should use a controlled keyword vocabulary to provide normalized discovery of data (expressed in various ways, e.g., thesauri, ontologies, etc.)

Categories of Discovery Level Metadata

What: Title of Data Set and Keywords Describing the Data Set

Why: Description and Purpose of the Data Set

When: Temporal Coverage of the Data Set

Who: Data Set Creator and Contact

Where: Geographic Extent and Location of Data Set Coverage

How: How the Data Set was Created and How to Access the Data

What does a metadata record look like?

- Human readable, e.g., as result from your favorite search tool (full record example of following screen shot at: http://mercury.ornl.gov/ornldaac/send/xsltText2;jsessionid=477B8A0DB8D7F17CBFFCD04CB0FB528F?fileURL=d%3A\mercury_instances\ornldaac\daac\harvested\record810.xml&full_datasource=DAAC+Datasets&full_queryString=+text+%3A+randerson+AND+%28+datasource+%3A%28+daac+daac++%29+%29+&ds_id=849)
- Machine readable: behind the scenes
  - As Text
  - As HTML
  - As XML

Human readable:
A Sample Search Result

DAAC Home -> MERCURY SEARCH

ORNL DAAC Distributed Active Archive Center for Biogeochemical Dynamics

About Us | About Data | Get Data | Data Tools | Help

Modify search

Metadata Report

Bookmark | Email | Help | Show Cart

Search Criteria: text : randerson and (datasource :(daac daac))

Title:	**GLOBAL FIRE EMISSIONS DATABASE, VERSION 2.1**
Project(s):	VEGETATION COLLECTIONS
Investigator(s):	COLLATZ, G.J. GIGLIO, L. KASIBHATLA, P.S. RANDERSON, J.T. VAN DER WERF, G.R.
Status:	Complete
Access Restrictions:	PUBLIC
Data Set Location:	The Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center (DAAC)
Data Center Contact:	ORNL DAAC User Services Office Oak Ridge National Laboratory Oak Ridge, Tennessee 37831 USA FAX: +1(865)574-4665 - ornldaac@ornl.gov Phone: +1(865)241-3952
Data Center URL:	http://daac.ornl.gov/
Data Set Citation:	Randerson, J. T., G. R. van der Werf, L. Giglio, G. J. Collatz, and P. S. Kasibhatla. 2007. Global Fire Emissions Database, Version 2.1. Data set. Available on-line [http://daac.ornl.gov/] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. doi:10.3334/ORNLDAAC/849
Download Data Sets:	GLOBAL FIRE EMISSIONS DATABASE, VERSION 2.1
Parameter Description:	

Parameter	Sensor	Source	Term	Topic
AIR TEMPERATURE	ANALYSIS	COMPUTER MODEL	ATMOSPHERIC TEMPERATURE	ATMOSPHERE
BIOMASS BURNING	ANALYSIS	COMPUTER	ENVIRONMENTAL IMPACTS	HUMAN

DAAC-recommended Citation

Machine readable:
A Sample Metadata Record – as text

co-ops.txt - WordPad

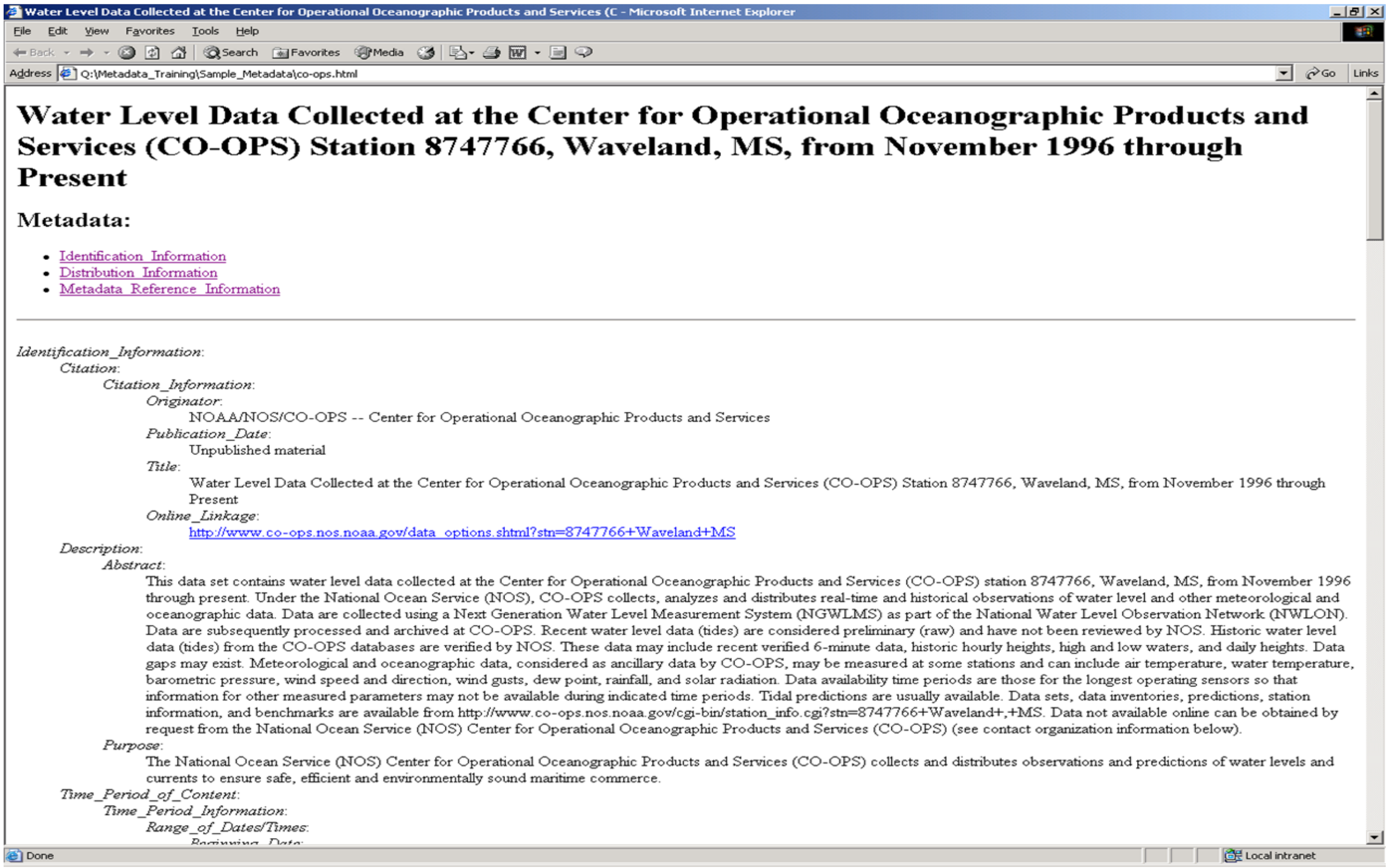
File Edit View Insert Format Help

Courier New | 10 | Western

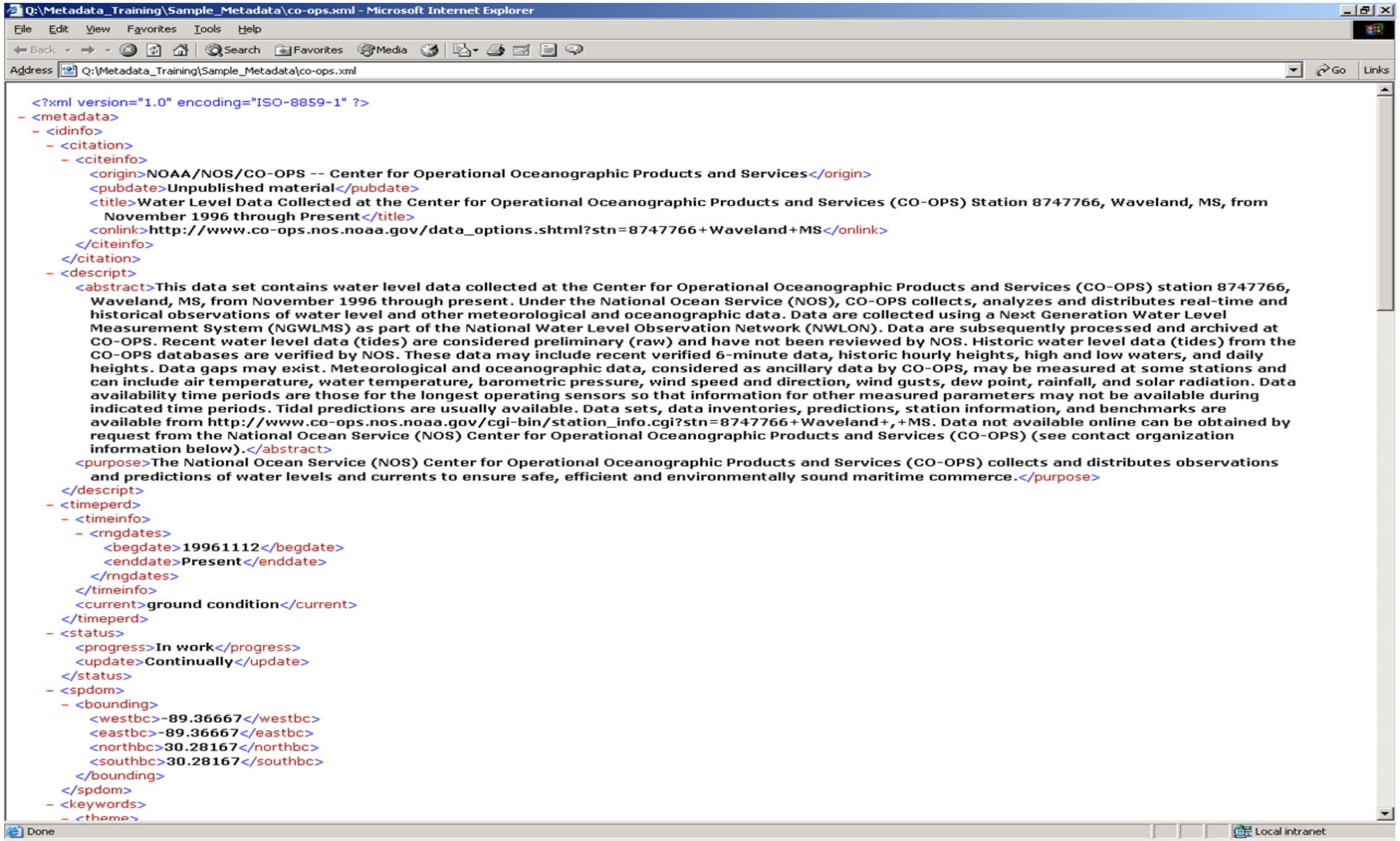
```
Identification_Information:
  Citation:
    Citation_Information:
      Originator: NOAA/NOS/CO-OPS -- Center for Operational Oceanographic Products and Services
      Publication_Date: Unpublished material
      Title: Water Level Data Collected at the Center for Operational Oceanographic Products and Services
(CO-OPS) Station 8747766, Waveland, MS, from November 1996 through Present
      Online_Linkage: http://www.co-ops.nos.noaa.gov/data_options.shtml?stn=8747766+Waveland+MS
  Description:
    Abstract: This data set contains water level data collected at the Center for Operational Oceanographic
Products and Services (CO-OPS) station 8747766, Waveland, MS, from November 1996 through present.  Under the
National Ocean Service (NOS), CO-OPS collects, analyzes and distributes real-time and historical
observations of water level and other meteorological and oceanographic data.
      Data are collected using a Next Generation Water Level Measurement System (NGWLMS) as part of the
National Water Level Observation Network (NWLON). Data are subsequently processed and archived at CO-OPS.
Recent water level data (tides) are considered preliminary (raw) and have not been reviewed by NOS.
Historic water level data (tides) from the CO-OPS databases are verified by NOS. These data may include
recent verified 6-minute data, historic hourly heights, high and low waters, and daily heights. Data gaps
may exist.  Meteorological and oceanographic data, considered as ancillary data by CO-OPS, may be measured
at some stations and can include air temperature, water temperature, barometric pressure, wind speed and
direction, wind gusts, dew point, rainfall, and solar radiation.  Data availability time periods are those
for the longest operating sensors so that information for other measured parameters may not be available
during indicated time periods.  Tidal predictions are usually available. Data sets, data inventories,
predictions, station information, and benchmarks are available from http://www.co-ops.nos.noaa.gov/cgi-
bin/station_info.cgi?stn=8747766+Waveland+,+MS. Data not available online can be obtained by request from
the National Ocean Service (NOS) Center for Operational Oceanographic Products and Services (CO-OPS) (see
contact organization information below).
    Purpose: The National Ocean Service (NOS) Center for Operational Oceanographic Products and Services
(CO-OPS) collects and distributes observations and predictions of water levels and currents to ensure safe,
efficient and environmentally sound maritime commerce.
  Time_Period_of_Content:
    Time_Period_Information:
      Range_of_Dates/Times:
```

For Help, press F1 NUM

Machine readable:
A Sample Metadata Record – as HTML



Machine readable:
A Sample Metadata Record – as XML



Preservation / Archiving metadata

- Why is it needed?
  - To support the long-term usability of digital data
  - To support the digital preservation process of a data archive or repository
  - To record information not covered by other metadata schemes that are useful for and/or collected by a data archive
- What is it? A work in process!
  - Many international and interdisciplinary efforts to define what is required in terms of metadata schemes from abstract to concrete:
    - OAIS Reference Model (OAIS RM) – generic
    - PREMIS: A generic approach coming from digital archives & libraries (based on OAIS RM)
    - FGDC & ISO 19115-2 (includes more than “preservation” metadata, but applied more directly to spatial & earth science data)

Categories of Preservation / Archiving Metadata based on OAIS RM

- For “content information” -- a set of information to be preserved over time –” aka “digital object” (DO)
- For a “designated community” or knowledge base
  - This is where content specifications and application profiles of metadata schemes come into play, e.g., NASA Content Specification for Earth Science data

- Preservation Description Information (PDI)
  - What is needed to efficiently manage & preserve a DO over time
- Representative Information (RI)
  - Best guesses as to what will be needed to be able to view or “render” the DO in future technical environments
  - Best guesses as to what will be needed for correct semantic interpretation of data
- Packaging Information
  - Means for binding DO including its MD into an identifiable unit for transfer and/or exchange
- Descriptive Information
  - Needed to facilitate efficient discovery & accessibility of preserved DO

Preservation / Archiving metadata – a work in progress!!

- Options for incorporating preservation MD into data set metadata
  - Analyses of existing metadata schemes to see if enough preservation MD is included
  - Analyses of whether / how data repository infrastructures accommodate preservation / archiving
- Need to TEST whether what is called for is truly needed, e.g.,
  - IMPORTANT NOT JUST TO **BACKUP**, BUT ALSO TO **RESTORE**!

- Recent analysis of what preservation MD is needed beyond ISO 19115-2
  - Shaon & Wolfe, D-Lib Magazine, Sept/Oct 2011.
- Comparison of PREMIS v1, FGDC & CIESIN's Geospatial Electronic Records for the Library of Congress' NGDA Project
  - Hoebelheinrich & Banning. 2008.
- Broad discussion of MD needs for preservation,
  - Duerr, et al
- *See* CIESIN's Geospatial Data Preservation Clearinghouse for more information (e.g., topic search on Preservation metadata for geospatial data)

Project Documentation

- Will vary by discipline and project
- Chance to use your

Imagination

about what could be useful to know in the future -- allowing for

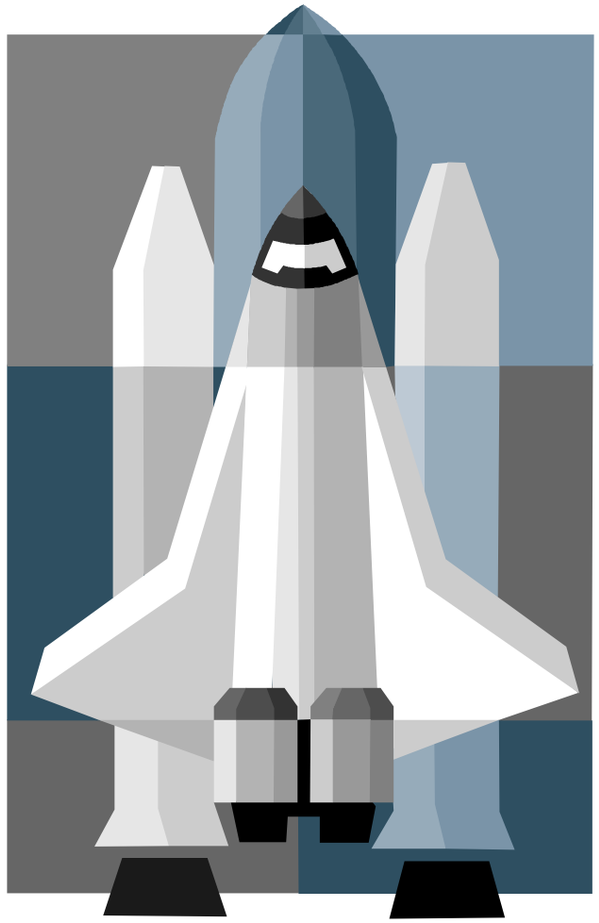
serendipity

& cyclical nature of scientific data

- Example from NASA Earth Science Data Preservation Content Specification (Nov 2011)
  - **3.3. Science Data Product Documentation content area**

Project Documentation

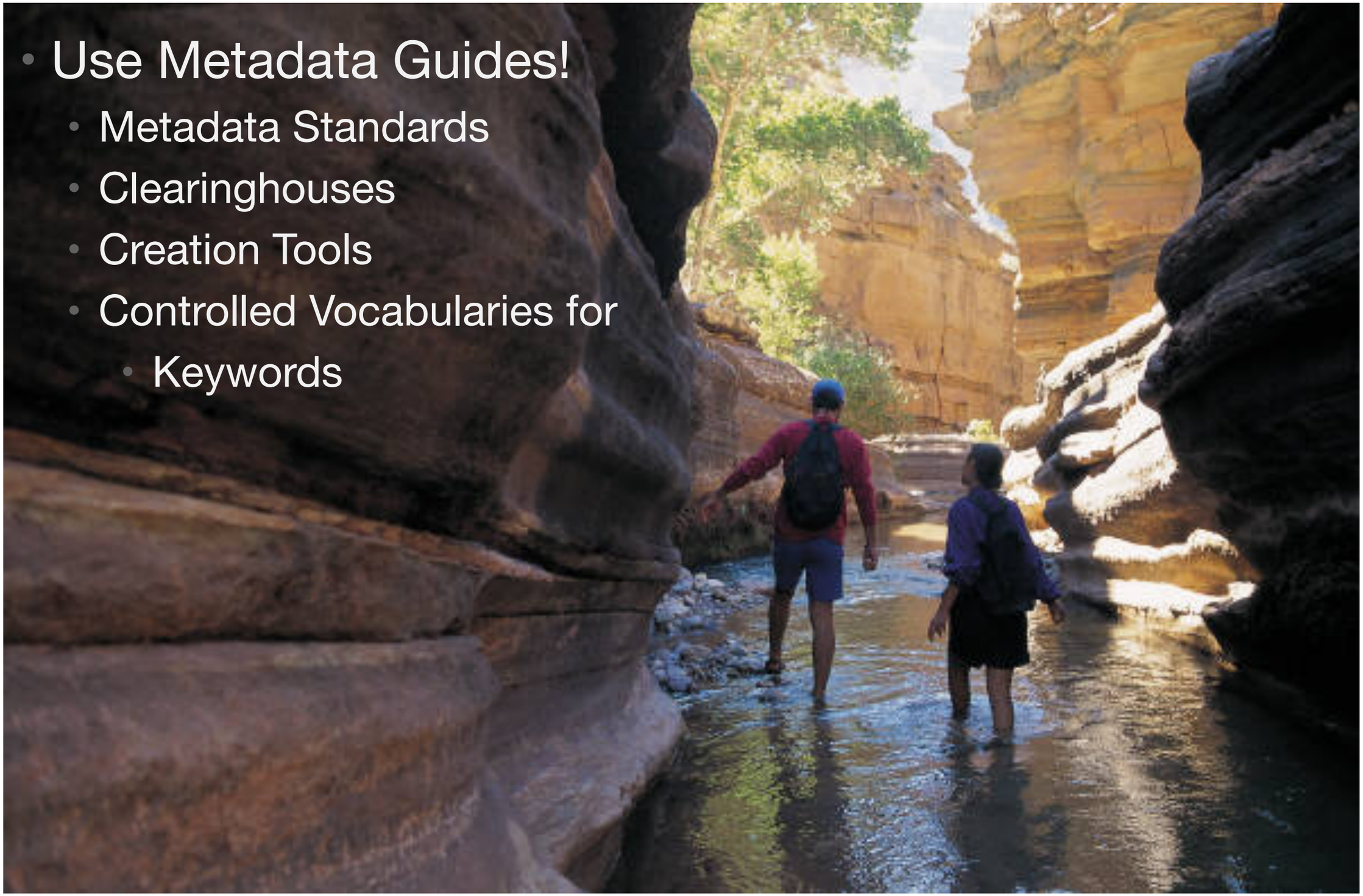
A Discipline Specific Example



Science Data Product Documentation

- **Categories of information that should be included with rationales for each:**
  - **Product Team 3.3.1**
  - **Product Requirements & Designs 3.3.2**
  - **Processing & Algorithm Version History 3.3.3**
  - **Product Generation Algorithms 3.3.4**
  - **Product Quality 3.3.5**
  - **Product Application 3.3.6**

Best Practices for Metadata Creation



What is a Metadata Standard?

- ***A Community agreed upon declaration that provides a structure to describe data with:***
- **Common terms to allow consistency between records**
- **Common definitions for easier interpretation**
- **Common language for ease of communication**
- **Common structure to quickly locate information**

- ***In search and retrieval, standards provide:***
- **A reliable and predictable format for computer interpretation**
- **A uniform summary description of the data set**

Many metadata standards exist

- **ISO 19115 Geographic information: Metadata**
  - Emphasis on geospatial data and services
- **World Meteorological Organization Core Metadata Profile (WMO) – a profile of ISO 19115**
  - Emphasis on meteorological data
- **Content Standard for Digital Geospatial Metadata (CSDGM)**
  - Federal Geographic Data Committee (FGDC)
  - Emphasis on geospatial data
- **Directory Interchange Format (DIF)**
  - Emphasis on Earth science data
- **Dublin Core Element Set**
  - Emphasis on web resources, publications

Two Standards: Two Sample Records

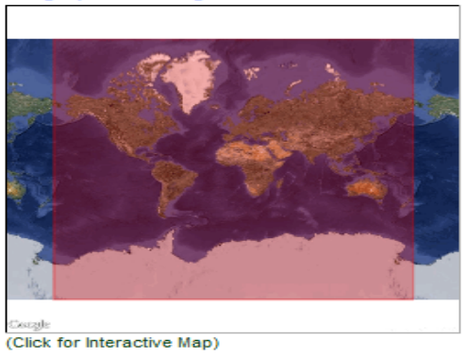
Directory Interchange Format (DIF)

Gridded Population of the World, Version 1 (GPWv1)
Entry ID: CIESIN_SEDAC_GPW_V1

[Update this Record]

Summary
Abstract: Gridded Population of the World, Version 1 (GPWv1) consists of estimates of human population for the year 1994 by 5 arc-minute grid cells. The 4 products are raw and smoothed population counts and density, all of which are available in several GIS-compatible data formats at the global and continent levels. A straight majority rule gridding algorithm, utilizing approximately 19,000 national and ... **Click to view more**

Geographic Coverage



(Click for Interactive Map)

Spatial coordinates
N: 90.0 **S:** -90.0 **E:** 180.0 **W:** -180.0

Data Set Citation
Dataset Creator: National Center for Geographic Information and Analysis (NCGIA); Center for International Earth Science Information Network (CIESIN)
Dataset Title: Gridded Population of the World, Version 1 (GPWv1)
Dataset Release Date: 1995
Dataset Release Place: Saginaw, MI
Dataset Publisher: CIESIN
Data Presentation Form: raster digital data
Online Resource: http://sedac.ciesin.columbia.edu/gpw/

Temporal Coverage
Start Date: 1994-07-01
Stop Date: 1994-07-01

Location Keywords
GEOGRAPHIC REGION > GLOBAL

Science Keywords
HUMAN DIMENSIONS > POPULATION > POPULATION DISTRIBUTION
HUMAN DIMENSIONS > POPULATION > POPULATION SIZE

ISO Topic Category
BOUNDARIES
SOCIETY

Project
EOSDIS > Earth Observing System Data Information System description
ESIP > Earth Science Information Partners Program description
GPW > Gridded Population of the World description
CWIC > CEOS WGISS Integrated Catalog description

Federal Geographic Data Committee

Identification_Information:
Citation:
Citation_Information:
Originator: LP DAAC
Publication_Date: 2006
Title: On Demand Digital Elevation Model V003
Edition:
Geospatial_Data_Presentation_Form: Remote Sensing Image
Series_Information:
Series_Name:
Issue_Identification:
Publication_Information:
Publication_Place: Sioux Falls, South Dakota, USA
Publisher: U.S. Geological Survey
Other_Citation_Details:
Online_Linkage:
Description:
Abstract:
The ASTER Digital Elevation Model (DEM) is a product that is generated from a pair of ASTER Level-1A images. This Level-1A input includes bands-3N (nadir)and -3B (aft-viewing) from the Visible Near Infra-Red telescope's along- track stereo data that is acquired in the spectral range of 0.78 to 0.86 microns. ASTER DEMs can be generated either with or without ground control points (GCPs). An Absolute DEM is created with GCPs that are supplied by an end-user who has requested the product. These DEMs have an absolute horizontal and vertical accuracy of up to 7 meters with appropriate GCPs and up to 10 meters without GCPs. Alternatively, a Relative DEM can also be generated without GCPs. These DEMs can be used to derive absolute slope and slope aspect which is good up to 5 degrees over a horizontal distance of over 100 meters. ASTER DEMs are expected to meet map accuracy standards for scales from 1:50,000 to 1:250,000. Data Set Characteristics: Area: ~60 km x 60 km Image Dimensions: 2500 rows x 2500 columns Input Image Resolution: 15 meters Output Image Resolution: 30 meters File Size: ~15 MB Data Type: 32-bit Vgroup Data Fields: 1 These level-3 data products were produced by the Land Processes DAAC (LP DAAC) from level-1 input supplied by the Ground Data System in Japan starting May 2, 2001, using early versions of the level-1 algorithm (03.00R02 and earlier).
Purpose:
Not Available
Supplemental_Information:
Not Available
Status:
Progress: Complete
Maintenance_and_Update_Frequency: As needed
Spatial_Domain:
Description_of_Geographic_Extent:
Bounding_Coordinates:
West_Bounding_Coordinate: -180.0
East_Bounding_Coordinate: 180.0
North_Bounding_Coordinate: 90.0
South_Bounding_Coordinate: -90.0
Keywords:
Theme:
Theme_Keyword_Thesaurus: GCMD SCIENCE PARAMETERS
Theme_Keyword_Thesaurus: GCMD PLATFORM
Theme_Keyword_Thesaurus: GCMD INSTRUMENT
Theme_Keyword_Thesaurus: PROJECT
Theme_Keyword_Thesaurus: ANCILLARY KEYWORDS
Theme_Keyword_Thesaurus: ISO TOPIC CATEGORY
Theme_Keyword_Thesaurus: DATA SET LANGUAGE
Theme_Keyword: EARTH SCIENCE > LAND SURFACE > TOPOGRAPHY > TERRAIN ELEVATION
Theme_Keyword: EARTH SCIENCE > LAND SURFACE > TOPOGRAPHY > TOPOGRAPHICAL RELIEF
Theme_Keyword: TERRA > EARTH OBSERVING SYSTEM, TERRA (AM-1)
Theme_Keyword: ASTER > ADVANCED SPACEBORNE THERMAL EMISSION AND

Metadata Creation Tools:

Some commercial & open source options

- MERMAid - Metadata Enterprise Resource Management Aid (National Coastal Data Development Center)
- ESRI ArcGIS tools
- EPA Metadata Editor (Environmental Protection Agency)
- FGDC Metadata Editor for ArcGIS (USGS)
- INSPIRE Geoportal Metadata Editor (European Commission)
- Global Change Master Directory's (GCMD) Docbuilder

Find out more from: http://www.fgdc.gov/metadata/geospatial-metadata-tools

For example, NOAA MERMAid

NOAA Satellite and Information Service
National Environmental Satellite, Data, and Information Service (NESDIS)

National Coastal Data Development Center

home » mermaid metadata resources » tools

Home | About Us | Metadata Resources | Regional Offices | Projects

training | *tools* | additional information | references

Metadata Enterprise Resource Management Aid (MERMAid)

NCDDC provides coastal data resources (organizations and individuals) with a tool to develop, validate, manage and publish metadata records via secure internet access. The Metadata Enterprise Resource Management Aid (MERMAid) allows users/data providers to establish unlimited metadata databases to organize their metadata records any way they see fit (i.e. by program, project, data type, personnel). Some of the key features in MERMAid include (1) user-defined roles and permissions at the metadata management and database levels; (2) change tracking; and (3) enhanced validation. Also, your existing FGDC compliant metadata (in XML format) can be ingested into and managed through MERMAid.

In the near future, NCDDC will be shifting from its current metadata catalog to a knowledge base catalog. MERMAid will play an integral role in this transition. To better leverage these new capabilities, enhanced search and discovery tools will be made available to the public and metadata managers that will provide powerful drill-down features.

Getting Started with MERMAid

- Request an Account
- V1.2 Getting Started Guide (PDF 5 MB)
- V1.2 Getting Started Guide

NCDDC Services
Site Map
Metadata Search
Middleware Technology
Download
Interactive Maps
Regional Ecosystems
C-SIDE

NOAA Services
Central Library
Photo Library
Video Library
Visualization Lab
Education Resources
National Locator
NOAA In Your State

SSC Visitor Services
Regional Map
Highway Map
Additional Information

More about MERMAid

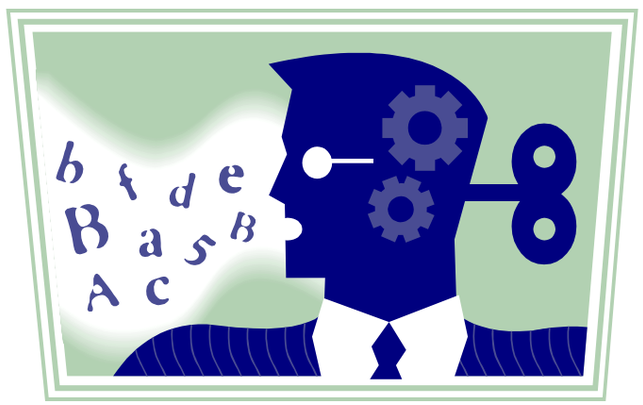
- Capabilities
  - Generates FGDC Standard, Biological, Shoreline, Remote Sensing Profile Records
  - Supports Ecological Metadata Language (EML)
  - Imports existing FGDC records in XML or .txt format
  - Exports records in XML, TXT, HTML or MARC XML via “save as”
  - Robust “Help” functions
- Technical Elements
  - Developed at National Coastal Data Development Center (NCDDC), uses the framework of the Z-Object Publishing Environment (ZOPE) as its web application server
  - Open source
  - Platform independent
  - Uses an object oriented dbase
  - Offers secure, remote access through a web browser

Find out more at: http://www.ncddc.noaa.gov/metadata-standards/mermaid/

Controlled Vocabularies for Keywords

- Keywords used to facilitate the identification & organization of descriptive information within and across different search systems
- Most useful to systems & searchers when they are:
  - Unambiguous
  - Specifically descriptive
  - Come from an established, published source based on continually updated use of discipline-specific language from formal & informal publications such as

    taxonomies, thesauri & ontologies



Useful sources for keywords

- Global Change Master Directory (GCMD)
  - Developed & maintained by National Aeronautics and Space Administration (NASA)
  - Focused upon Earth Science data sets & services
  - http://gcmd.nasa.gov/
- Getty Thesaurus of Geographic Names
  - Developed & maintained by The Getty Research Institute
  - Focused upon art, architecture & material culture, but definitive for established names of geographic names and places
  - http://www.getty.edu/research/tools/vocabularies/tgn/
- INSPIRE Feature Concept Dictionary
  - Developed & maintained by the European Environment Agency
  - Focused upon Spatial Data Themes
  - http://www.eionet.europa.eu/gemet/inspire_themes

Acknowledgements

- Robert Cook
  - Environmental Sciences Division, Oak Ridge National Laboratory
- Jeff Arnfield
  - National Climate Data Center, NOAA
- Viv Hutchison
  - US Geological Survey – NBII program
- Jaci Mize
  - National Coastal Data Development Center, NOAA
- Tyler Stevens
  - Wyle Information Systems (NASA Contractor)
- Ron Weaver
  - National Snow & Ice Data Center (NSIDC)

Questions?

References and Resources

- **Ball, C. A., G. Sherlock, and A. Brazma**. 2004. Funding high-throughput data sharing. Nature Biotechnology 22:1179-1183. doi:10.1038/nbt0904-1179.
- **Borer, ET., EW. Seabloom, M.B. Jones, and M. Schildhauer**. 2009. Some Simple Guidelines for Effective Data Management ,Bulletin of the Ecological Society of America. 90(2): 205- 214.
- **Christensen, S. W. and L. A. Hook**. 2007. NARSTO Data and Metadata Reporting Guidance. Provides reference tables of chemical, physical, and metadata variable names for atmospheric measurements. Available on-line at: http://cdiac.ornl.gov/programs/NARSTO/
- **Cook, Robert B, Richard J. Olson, Paul Kanciruk, and Leslie A. Hook**. 2001. Best Practices for Preparing Ecological Data Sets to Share and Archive. Bulletin of the Ecological Society of America, Vol. 82, No. 2, April 2001.
- **Duerr et al.** "Challenges in Long-Term Data Stewardship",http://storageconference.org/2004/Papers/05-Duerr-a.pdf
- **Federal Geographic Data Committee (FGDC).** Geospatial Metadata. http://www.fgdc.gov/metadata
- **Geospatial Data Preservation Clearinghouse. CIESEN at Columbia University.** Search for Preservation Metadata: http://geopreservation.org/topicResults.jsp?types=8
- **Getty Thesaurus of Geographic Names**. http://www.getty.edu/research/tools/vocabularies/tgn/
- **Hoebelheinrich & Banning.** "An Investigation into Metadata for Long-Lived Geospatial Data Formats", 2008. http://www.ngda.org/reports/InvestigateGeoDataFinal_v2.pdf
- **Hook, L. A., T. W. Beaty, S. Santhana-Vannan, L. Baskaran, and R. B. Cook**. June 2007. Best Practices for Preparing Environmental Data Sets to Share and Archive. http://daac.ornl.gov/PI/pi_info.shtml
- **INSPIRE Feature Concept Dictionary.** http://www.eionet.europa.eu/gemet/inspire_themes

References and Resources (continued)

- **Kanciruk, P., R.J. Olson, and R.A. McCord**. 1986. Quality Control in Research Databases: The US Environmental Protection Agency National Surface Water Survey Experience. In: W.K. Michener (ed.). Research Data Management in the Ecological Sciences. The Belle W. Baruch Library in Marine Science, No. 16, 193-207.
- **Management in the Ecological Sciences.** The Belle W. Baruch Library in Marine Science, No. 16, 193-207.
- **Michener, W K.** 2006. Meta-information concepts for ecological data management. Ecological Informatics. 1:3-7.
- **Michener, W.K. and J.W. Brunt (ed.).** 2000. Ecological Data: Design, Management and Processing, Methods in Ecology, Blackwell Science. 180p.
- **Michener, W. K., J. W. Brunt, J. Helly, T. B. Kirchner, and S. G. Stafford**. 1997. Non-Geospatial Metadata for Ecology. Ecological Applications. 7:330-342.
- **NASA Earth Science Data Preservation Content Specification (Nov 2011).** http://earthdata.nasa.gov/sites/default/files/field/document/NASA_ESD_Preservation_Spec.pdf
- **NASA, 2011: Metadata Requirements – Base Reference for NASA Earth Science Data Products, (Nov 2011).** http://earthdata.nasa.gov/sites/default/files/field/document/NASA%20ESD%20Base%20Metadata%20Requirements_V1_20110922_0.pdf
- **National Aeronautics and Space Administration (NASA). Global Change Master Directory.** http://gcmd.nasa.gov.

References and Resources (continued)

- **National Research Council (NRC).** 1991. Solving the Global Change Puzzle: A U.S. Strategy for Managing Data and Information. Report by the Committee on Geophysical Data of the National Research Council Commission on Geosciences, Environment and Resources. National Academy Press, Washington, D.C.
- **NOAA GEO-IDE best practices wiki**. https://geo-ide.noaa.gov/wiki/index.php?title=ISO_Lineage
- **Preservation Metadata (PREMIS).** http://www.loc.gov/standards/premis/
- **Shaon & Woolfe.** "Long-term Preservation for Spatial Data Infrastructures: a Metadata Framework and Geo-portal Implementation", D-Lib Magazine, September/October 2011. doi:10.1045/september2011-shaon. http://www.dlib.org/dlib/september11/shaon/09shaon.print.html
- **Taylor, Mark. 2000.** "Developing Spatial Data Infrastructures – The SDI Cookbook". Global Spatial Data Infrastructure Association. http://www.gsdi.org/pubs/cookbook/chapter03a.html
- **U.S. EPA. 2007. Environmental Protection Agency Substance Registry System (SRS).** SRS provides information on substances and organisms and how they are represented in the EPA information systems. Available on-line at: http://www.epa.gov/srs/
- **USGS. 2000.** Metadata in plain language. Available on-line at: http://geology.usgs.gov/tools/metadata/tools/doc/ctc/